J. Med. Microbiol. — Vol. 47 (1998), 893–898

EPIDEMIOLOGY OF INFECTION

Streptococcal *emm* types associated with T-agglutination types and the use of conserved *emm* gene restriction fragment patterns for subtyping group A streptococci

B. BEALL, R. R. FACKLAM, J. A. ELLIOTT, A. R. FRANKLIN, T. HOENES, D. JACKSON, L. LACLAIRE, T. THOMPSON and R. VISWANATHAN

Centers for Disease Control and Prevention, Childhood and Respiratory Infections, Atlanta, GA 30333, USA

The T-agglutination types were determined for a diverse collection of 1531 group A streptococci for which the 5′ M protein gene (*emm*) sequences had been analysed. The majority of the T-agglutination types correlated with previously seen M/*emm*/T-type associations; however, several new associations were found. Analysis of a subset of this collection – which included 1157 clinical isolates with multiply encountered *emm* types – found that *emm* amplicon restriction profiles of isolates sharing identical T types and opacity factor phenotypes are useful for detecting groups of isolates with identical *emm* genes. Many *emm* genes of known 5′ sequence display a highly conserved restriction pattern amongst clinical isolates widely separated both geographically and temporally.

Introduction

Subtyping of clinical isolates of group A streptococci (GAS) has relied historically upon serotyping of the surface M protein with polyclonal antisera. However, there are many problems inherent to M typing, including the large number of M serotypes and the lack of availability of comprehensive M typing sera sets. Currently, there are *c.* 80 recognised M serotypes, 15 provisional M types, and many more deduced new specificities on the basis of *emm* gene sequences that encode the hypervariable regions of M proteins. A recent study showed that rapid *emm* variable region sequence analysis (*emm* typing) of GAS in combination with conventional T-agglutination pattern determination and opacity factor (OF) phenotype determination, is a useful means by which to subtype this organism. This system is almost 100% efficient and often deduces the M serotype [1, 2]. This facet of *emm* typing is especially important because *emm* gene sequences correlated with known serospecificities are accessible from the GenBank and for the most part complement *c.* 50 years of accumulated M serotyping data [1–10]. The present study shows the T-agglutination types associated with the various *emm* types that were encountered over the last 3 years. Furthermore, an improvement of the *emm* typing system that takes advantage of *emm* type specific conserved restriction patterns is clearly demonstrated.

Received 8 Oct. 1997; accepted 2 Feb. 1998.
Corresponding author: Dr B. Beall.

Materials and methods

Strains

Approximately 64% of the 1531 strains for which T-agglutination patterns were determined represented the total number of sterile site isolates obtained from five regions within the USA for 1–3 years during the period 1994–1997. These regions included metropolitan Atlanta, GA; metropolitan San Francisco, CA; metropolitan Minneapolis, MN; Connecticut state and Oregon state. Approximately 13% of the strains represented sporadic sterile site isolates from other regions within the USA. The remaining strains were obtained from various sources outside the USA including Bulgaria, Poland, Malaysia, Korea, Australia, New Zealand and South America. Approximately 100 strains were M type and M provisional type CDC reference strains.

Subtyping procedures

T typing and OF determination were performed as described previously [8, 11]; *emm*-specific PCR products were prepared and the 5′ end of *emm* genes were sequenced for type determination as described pre-

viously [2]. For generation of an *emm* amplicon restriction profile (ERP), 8.5 μl of unpurified PCR product were digested in a total volume of 10 μl containing 1× reaction buffer 4 (Gibco-BRL), 5 units of *Hae* III and 5 units of *Hinc* II. The samples were incubated for 1 h (or longer) at 37°C, electrophoresed in agarose 2% w/v gels containing ethidium bromide 1 μg/ml in TAE buffer (40 mM Tris-acetate, pH 8.2, 20 mM sodium acetate, 1 mM EDTA) at 100 V for 30 min, visualised on a UV transilluminator and photographed.

Sequences were given the indicated GenBank *emm* type designations if ≥ 95% identical over at least the first 160 bases, permitting one in-frame deletion or insertion of up to seven codons or one frameshift of up to seven codons. Such derivatives made up < 3% of the total sequences represented. GenBank accession numbers for most of the *emm* type sequences are given in references [1, 2, 5, 9, 10, 12, 13 and 14] except for the *emm* sequences found from our M67, M68, and M69 reference strains (*emm67*, *emm68*, and *emm69* in Table 1) that differed from the data described by Whatmore and colleagues [10]. Our M65 and M69 reference strains have the same 5′ *emm* sequence, which was identical to the previously determined *emm65* sequence [10]. It has also been determined that M44 and M61 strains have the same 5′ *emm* sequence [1, 10]. The sequences shown above that are not listed in these references were obtained in this laboratory and the accession numbers are as follows: *emm67* = AF025949, *emm68* = AF025948, ST2034 = U74320, PT2631 = U94589, ST4935 = U92492, *emm70* = AF035838.

Results

T-agglutination patterns seen in different emm types

Table 1 lists the T patterns and OF reactions that have been determined for 90 *emm* types. Data are not given for the 25 new *emm* types that were found only amongst sporadic isolates. The majority (1244 of 1363 T-typable isolates) of the T agglutination patterns shown in Table 1 correlated with previously published M type/T type/OF associations [1–5], but there were also many unpublished associations (Table 1). Nearly 12% of the isolates (178/1531) were found to be T non-typable.

emm specific amplicon restriction profiling

Table 2 shows that 945 of 1157 geographically diverse clinical isolates had a predominant ERP type shared within their respective *emm* sequence types. These isolates were collected primarily from normally sterile sites, although *c.* 2% of them were isolated from throat or skin lesion cultures. These isolates represented 32 different *emm* sequence types that we have encountered on multiple occasions. For each of the *emm* types with the exception of *emm6*, *emm76* and ST2974, there was a single restriction pattern accounting for > 60% of isolates within that *emm* type; however, for some of the

Table 1. *emm* types associated with T-agglutination patterns and opacity factor (OF) reaction

T pattern, OF	*emm* type* (number of isolates examined)
T1, OF−	**1**(258), 28†(2)
T1, OF+	**68**(2)
T3/13/B3264, OF−	**3**(108), **33**(19), **43**(14), STNS5‡(13), **53**(12), **PT2110**(5), **42**(4), 80†(3), ST88/25‡(4), **65/69**(3), **41**(2), ST90/85‡(2), 74†(3), 67†(1), 64‡(1), **56**(1), **52**(1), **39**(1) 34‡(1), 1(1)
T3/13/B3264, OF+	**13**(24), **77**(22), **73**(21), **PT2612**(16), PT4245‡(7), 68‡(4), ST2034‡(4), **81**(1), 75‡(1), **73**(1), 68‡(1), 64‡(1), 22(1), **PT4931**(1)
T2, OF+	**2**(14), 4(3), 6(1), 58(1), 75†(1), STNS14X‡(3)
T4, OF+	**4**(35), **60**(7), **63**(1), 58‡(1), 22‡(1), **46**(1)
T6, OF−	**6**(28), **ST64/14**(8), 70‡(1)
T28, OF+	**28**(108), **PT2841**(31), **77**(9), 2(2), 4(1), **48**(1)
T28, OF−	**56**(2), 71†(1), **70**(1), STNS1‡(1)
T11/12, OF+	**11**(56), **PT4245**(30), **22**(37), **59**(16), 66(13), ST2967‡(12), **76**(11), **78**(8), **44/61**§(7), **62**(6), 73‡(1), PT2841‡(2), 80‡(1)
T11/12, OF−	**12**(130), STNS14X‡(2), 72(1), 68‡(1)
T5/27/44, OF+	**PT180**(21), **27**(13), **44/61**(7), 76(1), 11(1), 8‡(1)
T5/27/44, OF−	**5**(24)
T14, OF−	**PT2631**‡(18), **80**(2), **14**(2), **51**(1)
T14, OF+	**49**(13), ST2967‡(10), 9‡(4), **59**(2), 60‡(1), 64‡(2), 76‡(1)
T8/25/IMP19, OF+	**75**(28), **ST2974**(23), **58**(22), 76‡(9), **8**(8), 63‡(4), **65**(2), **25**(2), **2**(2), 4(2), ST2967‡(2), 80‡(1), **79**(1), **65/69**(1), 59(1), **57**(1), **55**(1), **PT3875**(1), **PT2233**(1)
T9, OF+	**9**(3)
T9, OF−	**74**(2), **18**(1), **36**(1), **67**(1)
T23/8/14, OF−	**15**(1), **17**(1), **23**(1), **32**(1), **47**(1)
T22, OF+	**22**(2), 76‡(3)
T non-typable, OF+	ST2034(15), ST2967(13), PT4245(9), 59(8), 4(8), ST4935(7), 65/69(5), 22(5), 11(5), 25(4), PT3875(4), PT2612(3), 81(2), 75(2), 63(2), PT180(2), 78(1), 77(1), 76(1), 66(1), 49(1), 28(1), 13(1), 2(1)
T non-typable, OF−	6(13), 3(11), 18(7), 56(5), 43(4), 12(4), 33(3), 1(3), 80(2), 53(2), 41(2), 31(2), ST64/14(2), STNS5(2), 67(1), 54(1), 39(1), 37(1), 32(1), 30(1), 29(1), 26(1), 24(1), 19(1), 5(2), PT2631(1), STNS14X(1)

**emm* types expected for the T pattern and OF reaction are in bold.
†Represents an unpublished M (*emm*)/T-agglutination pattern association.
‡M (*emm*) type/T-agglutination pattern associations have not been previously made for this *emm* type.

Table 2. Major *emm* amplicon restriction profiles shared among isolates of the same *emm* sequence type

emm type	Number (%) of isolates	Number (%) from the USA	Number (%) from other countries
emm1	225 (99)	198 (88)	27 (12): S. America
emm2	17 (100)	12 (71)	5 (29): S. America, Bulgaria
emm3	104 (98)	81 (78)	23 (22): S. America, Korea, Bulgaria
emm4	34 (79)	27 (79)	7 (21): S. America, Poland, Malaysia
emm5	14 (61)	13 (93)	1 (7): Malaysia
emm6	17 (50)	9 (53)	8 (47): S. America, Korea, Poland
emm11	34 (87)	33 (97)	1 (3): S. America
emm12	120 (100)	95 (79)	25 (21): S. America, Bulgaria, Malaysia
emm13	10 (83)	6 (60)	4 (40): S. America
emm18	6 (100)		
emm22	28 (93)	19 (68)	9 (32): S. America
emm28	94 (97)	89 (95)	5 (5): Korea, Poland
emm33	28 (100)	27 (96)	1 (4): Korea
emm41	4 (67)	4 (100)	
emm43	12 (100)	12 (100)	
emm44/61	7 (100)	7 (100)	
emm49	5 (100)	4 (80)	1 (20): S. America
emm53	7 (100)	6 (86)	1 (14): New Zealand
emm58	11 (65)	11 (100)	
emm59	17 (100)	16 (88)	1 (12): S. America
emm65/69	8 (100)		8 (100): Bulgaria, S. America
emm66	8 (100)	8 (100)	
emm73	4 (100)	4 (100)	
emm75	18 (100)	12 (67)	6 (33): Korea, S. America
emm76	7 (34)	7 (100)	
emm77	33 (97)	26 (79)	7 (21): S. America
PT4245	35 (100)	34 (97)	1 (3): S. America
PT2841	17 (74)	1 (6)	16 (94): S. America
PT2110	3 (100)	3 (100)	
PT180	10 (83)	6 (60)	4 (40): S. America, Malaysia
2974-1	8 (57)	5 (62)	3 (38): S. America

*Percentage of isolates with the most frequently occurring *emm* restriction profile for the indicated *emm* type.

emm types the results could be biased because of small sample size or geographic distribution. Twenty-four of these major profiles are shown in Fig. 1 (lanes 2, 3, 5–14, 16–29), and for *emm6* two secondary ERPs are shown (Fig. 1, lanes 1 and 4). As the *emm* gene amplicons of different types vary extensively in size (*c.* 900–1600 bp) and usually have two or more *Hae* II or *Hinc* II sites, or both, in most cases major restriction profiles representing one *emm* type are readily distinguishable from the major ERP of other *emm* types.

A minority of *emm* genes lacked both *Hinc* II and *Hae* III sites (lanes 12, 19 and 21 in Fig. 1); they were distinguishable from other *emm* genes lacking these sites by differing sizes (Fig. 1; compare amplicons for *emm33* and *emm43* in lanes 19 and 21 with *emm66* amplicon in lane 12) or by unique restriction profiles generated by other frequent cutters such as *Rsa* I or *Dpn* II. Although the *emm33* and *emm43* PCR fragments lack *Hinc* II and *Hae* III sites, and were almost of identical size (Fig. 1, lanes 19 and 21), the two PCR fragments were readily distinguishable by *Rsa* I cleavage profiles. Regardless of amplicon sizes, alternate restriction enzymes are always used for amplicons lacking *Hinc* II and *Hae* III sites.

Currently, for GAS subtyping ERPs are generated in association with *emm* sequence results at this laboratory as follows; isolates are subjected to T-antigen typing and OF determination as described previously [8, 11]. Those isolates from the same general population area during a 1-year time period that have related T-antigen patterns, the same OF phenotype and identical ERP are identified. A random sample of *emm* amplicons from such a group is subjected to *emm* variable region sequence analysis, which effectively determines the *emm* type of the entire group. A false deduced *emm* designation result has not been found with this scheme.

An example of the routine use of ERP comparison to reduce the numbers of samples sequenced is shown in Fig. 2. Lanes 2–10 represent the ERPs of nine T type 28, OF+ Brazilian isolates. Lanes 2 and 5–10 are the most common *emmPT2841* ERP (compare with Fig. 1, lane 8). *emm* sequence analysis of two of these isolates showed almost 100% identity to the *emm PT2841* gene. Lane 3 had a different ERP; however, this PCR fragment was also found to have the *emmPT2841* sequence. Comparison of these two *emmPT2841* ERPs indicated the probable loss of one *Hinc* II or *Hae* III site in the amplicon shown in

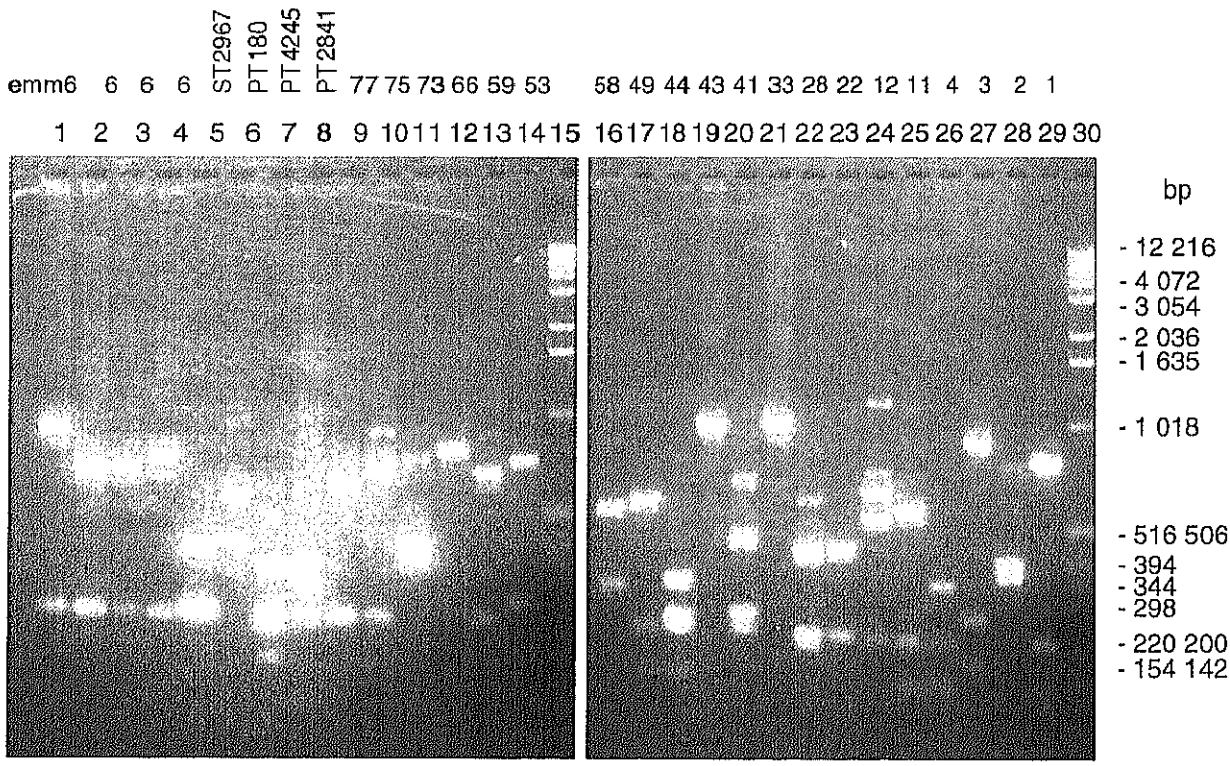

Fig. 1. Frequent *emm* amplicon restriction profiles (*Hinc* II and *Hae* III double digest profiles) of 25 different *emm* types. Lanes 1–4, depict three *emm6* profiles; 5–14, the most frequently encountered *emm* amplicon restriction profiles for the indicated *emm* types; 15 and 30, Gibco-BRL kb ladder size standard.

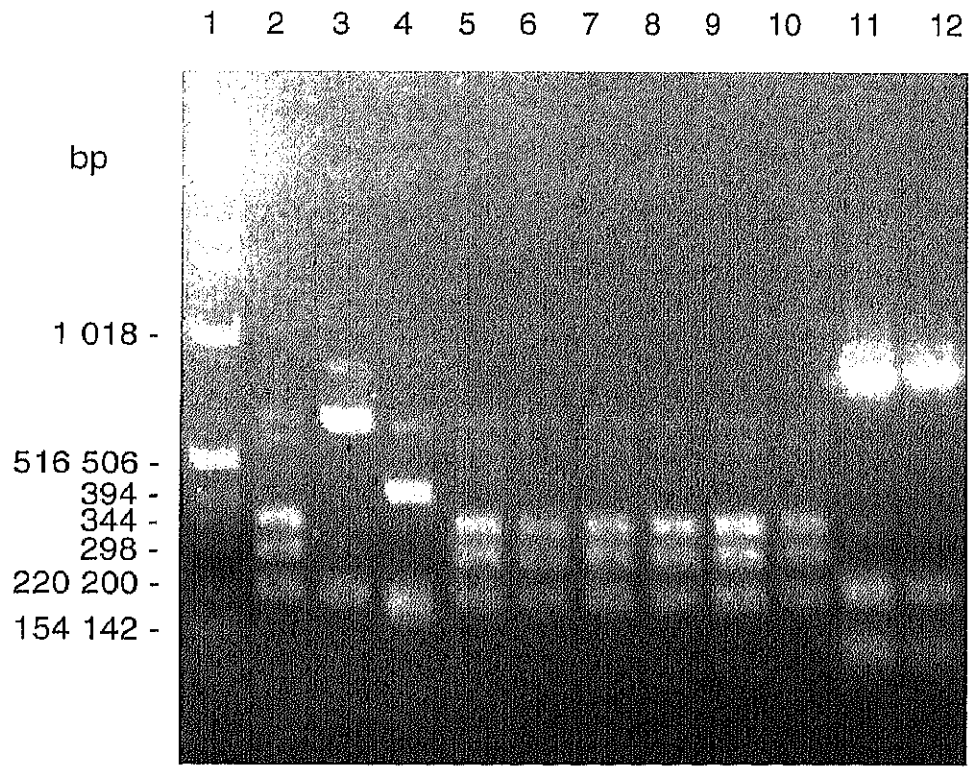

Fig. 2. ERP analysis of 11 Brazilian GAS isolates. Lanes 2–10, were obtained from nine independent T28 OF+ isolates; 11 and 12, were obtained from two independent T1 OF− isolates. Selected uncut samples were subjected to sequence analysis and found to be PT2841 (lanes 2, 3 and 5), *emm28* (4) and *emm1* (11).

lane 3. One of the T28 OF+ isolates produced the most common *emm28* ERP (compare Fig 2, lane 4 with Fig. 1, lane 22) and sequence analysis revealed near identity with *emm28* (not shown). Also shown in Fig. 2 is the major *emm1* ERP (Fig. 1, lanes 11 and 12) shared by two independent T type 1, OF−

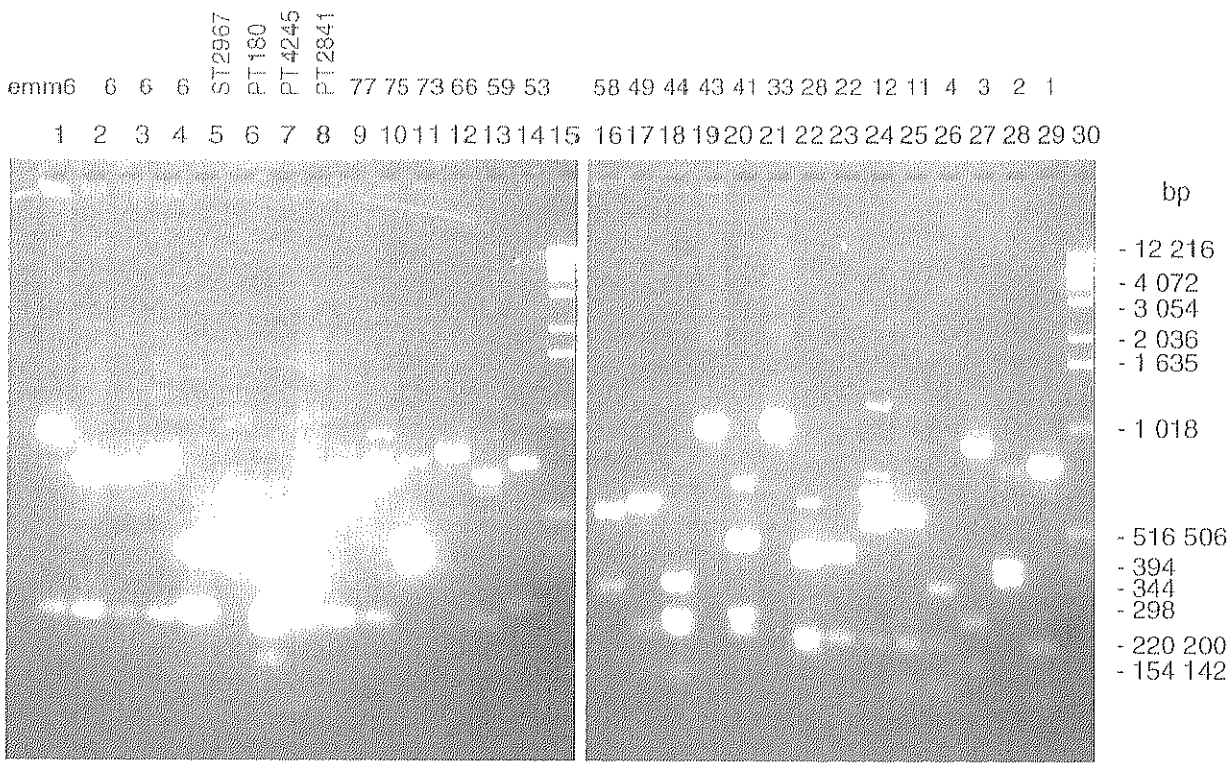

Fig. 1. Frequent *emm* amplicon restriction profiles (*Hinc*II and *Hae*III double digest profiles) of 25 different *emm* types. Lanes 1–4, depict three *emm6* profiles; 5–14, the most frequently encountered *emm* amplicon restriction profiles for the indicated *emm* types; 15 and 30, Gibco-BRL kb ladder size standard.

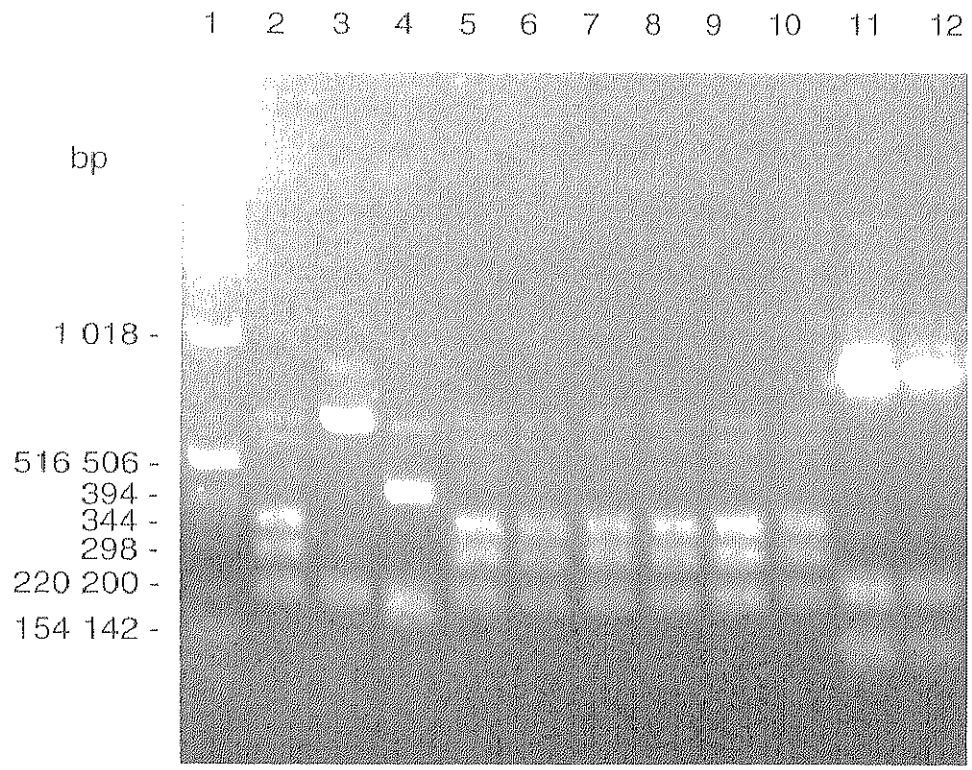

Fig. 2. ERP analysis of 11 Brazilian GAS isolates. Lanes 2–10, were obtained from nine independent T28 OF+ isolates; 11 and 12, were obtained from two independent T1 OF− isolates. Selected uncut samples were subjected to sequence analysis and found to be PT2841 (lanes 2, 3 and 5), *emm28* (4) and *emm1* (11).

lane 3. One of the T28 OF+ isolates produced the most common *emm28* ERP (compare Fig 2, lane 4 with Fig. 1, lane 22) and sequence analysis revealed near identity with *emm28* (not shown). Also shown in Fig. 2 is the major *emm1* ERP (Fig. 1, lanes 11 and 12) shared by two independent T type 1, OF−

isolates. Sequence analysis of one of the isolates revealed identity to the *emm1* sequence.

ERP analysis is useful for deducing the *emm* type for epidemiologically related sets of isolates that contain variant alleles of previously characterised *emm* genes. For example, two *emm3* isolates with ERP profiles not typical of the major *emm3* ERP were four codon deletion derivatives of *emm3*. In addition, on numerous occasions ERP analysis has been used for the identification of epidemiologically related sets of isolates containing *emm* genes that have not been encountered previously. For such isolates, that have related T patterns, identical OF phenotypes and identical ERP, it is essential to examine the 5′ variable *emm* sequence of only one of the isolates to predict the *emm* type of the whole group (unpublished data).

Discussion

The sample set in the present study was biased in that primarily isolates from the USA were analysed. Even so, it is interesting that relatively few new *emm* sequence types have been encountered. The definition of an *emm* type is given here as ⩾ 95% identical over the first 160 bases of sequence obtained with primer emmseq2 [2], permitting one in-frame deletion or insertion of up to seven codons or one frameshift of up to seven codons. The majority of the known *emm* types are only 50–85% identical over the first 160 bases obtained. The basis of the relative rarity of 'intermediate' *emm* types with 85–95% sequence identity has not yet been determined.

For most of the *emm* genes analysed the study found a striking lack of allelic diversity determined by sequence and ERP analysis from geographically distinct isolates (Table 1). These results may indicate that these are evolutionarily stable *emm* genes that infrequently undergo genetic variation. In contrast, for *emm* genes such as *emm6*, *emm76* and ST2974 multiple alleles of different sizes were encountered (three of *emm6* are shown in Fig. 1, lanes 1–4), at least partially due to different patterns of homologous recombination between direct repeats in the hypervariable region [15, unpublished observations].

For some isolates with the same *emm* types and ERPs more than one T agglutination pattern is encountered. We have found no correlation between different ERP subtypes within a given *emm* type with differences in T types. Generally, we have found that pulsed-field gel electrophoresis patterns of chromosomal digests reflect that isolates of the same *emm* type, but with different T types, are genetically divergent (unpublished observations), whilst isolates with identical T and *emm* types are more closely related. The compilation of the varied *emm*/T-antigen pattern associations shown in Table 1 is an indication of the genetic diversity within *emm* types that is due to horizontal genetic transfer events between unrelated strains.

The results shown here, which show major conserved *emm* restriction fragment patterns for 32 *emm* types, are analogous to those found previously when *Hae* III was used to find characteristic *emm* restriction profiles within three different GAS serotypes [16]. Furthermore, outbreak-related strains could be distinguished from other strains within the same M-type by the use of characteristic *Hae* III profiles [16]. We have also found this strategy useful for this purpose (unpublished data); however, the combination of *Hae* III and *Hinc* II often provides more discriminatory electrophoretic profiles than that of *Hae* III alone.

For many *emm* types there is a high probability of deducing *emm* types from geographically diverse isolates by their ERP alone in combination with T typing and OF determination. For example, this laboratory has rarely encountered alternate ERPs for the distinctive patterns seen for the commonly encountered *emm1* and *emm3* types, and has not found an alternative ERP for the frequently occurring *emm12* isolates (Table 1 and Fig. 1, lanes 24, 27 and 29). For commonly encountered ERP profiles, *emm* types can be detected by concurrently electrophoresing digest standards or comparison with saved electrophoretic patterns on a computer scanner.

We thank the Emerging Infections Programs (EIP), supported by the National Center for Infectious Diseases for isolates resulting from active surveillance; California EIP: A. Reingold, Gretchen Rothrock and D. Vugia; Connecticut EIP: M. Carter, J. Hadler, Patricia Mshar, C. Morin and Nancy Barrett; Georgia EIP: Wendy Baughman and Monica Farley; Minnesota EIP: Jean Rainbow, Kristine MacDonald and M. Osterholm; Oregon EIP: S. Ladd-Wilson, Karen Stefonek, P. Cieslak; and Katherine O'Brien for isolates from sporadic cases of necrotising fasciitis.

References

1. Beall B, Facklam R, Thompson T. Sequencing *emm*-specific PCR products for routine and accurate typing of group A streptococci. *J Clin Microbiol* 1995; 34: 953–958.
2. Beall B, Facklam R, Hoenes T, Schwartz B. A survey of emm gene sequences and T-antigen types from systemic *Streptococcus pyogenes* infection isolates collected in San Francisco, California; Atlanta, Georgia, and Connecticut in 1994 and 1995. *J Clin Microbiol* 1997; 35: 1231–1235.
3. Colman G, Tanna A, Efstratiou A, Gaworzewska ET. The serotypes of *Streptococcus pyogenes* present in Britain during 1980–1990 and their association with disease. *J Med Microbiol* 1993; 39: 165–178.
4. Fraser CAM, Colman G. Some provisional M-types among *Streptococcus pyogenes* (Lancefield group A). In: Kimura Y, Kotami S, Shiokawa Y (eds) Recent advances in streptococci and streptococcal diseases. Bracknell, Reedbooks. 1985: 35–36.
5. Johnson DR, Kaplan EL. A review of the correlation of T-agglutination patterns and M-protein typing and opacity factor production in the identification of group A streptococci. *J Med Microbiol* 1993; 38: 311–315.
6. Kaufhold A, Podbielski A, Baumgarten G, Tap J, Blokpoel M, Schouls L. Rapid typing of group A streptococci by the use of DNA amplification and non-radioactive allele-specific oligonucleotide probes. *FEMS Microbiol Lett* 1994; 119: 19–26.

7. Lancefield RC. Current knowledge of the type-specific M antigens of group A streptococci. *J Immunol* 1962; **89**: 307–313.
8. Moody MD, Padula J, Lizana D, Hall CT. Epidemiologic characterization of group A streptococci by T-agglutination and M-precipitation tests in the public health laboratory. *Health Lab Sci* 1996; **2**: 149–162.
9. Podbielski A, Melzer B, Lutticken E. Application of the polymerase chain reaction to study the M protein(-like) gene family in beta-hemolytic streptococci. *Med Microbiol Immunol* 1991; **180**: 213–227.
10. Whatmore AM, Kapur V, Sullivan DJ, Musser JM, Kehoe MA. Non-congruent relationships between variation in *emm* gene sequences and the population genetic structure of group A streptococci. *Mol Microbiol* 1994; **14**: 619–631.
11. Maxted WR, Widdowson JP, Fraser CAM, Ball LC, Bassett DCJ. The use of the serum opacity reaction in the typing of group-A streptococci. *J Med Microbiol* 1973; **6**: 83–90.
12. Pack TD, Podbielski A, Boyle MDP. Identification of an amino acid signature sequence predictive of protein G-inhibitable IgG_3-binding activity in group-A streptococcal IgG-binding proteins. *Gene* 1996; **171**: 65–70.
13. Relf WA, Martin DR, Sriprakash KS. Identification of sequence types among the M-nontypeable group A streptococci. *J Clin Microbiol* 1992; **30**: 3190–3194.
14. Relf WA, Martin DR, Sriprakash KS. Antigenic diversity within a family of M proteins from group A streptococci: evidence for the role of frameshift and compensatory mutations. *Gene* 1994; **144**: 25–30.
15. Hollingshead SK, Fischetti VA, Scott JR. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol Gen Genet* 1987; **207**: 196–203.
16. Stanley J, Desai M, Xerry J, Tanna A, Efstratiou A, George R. High-resolution genotyping elucidates the epidemiology of group A *Streptococcus* outbreaks. *J Infect Dis* 1996; **174**: 500–506.